

Can a Multimodel SuperEnsemble technique be used for precipitation forecasts?

D. Cane and M. Milelli

Regional Environmental Protection Agency – Arpa Piemonte, Torino, Italy

Received: 1 October 2009 – Revised: 19 February 2010 – Accepted: 19 February 2010 – Published: 8 March 2010

Abstract. The Multimodel SuperEnsemble technique is a postprocessing method for the estimation of weather forecast parameters reducing direct model output errors. It differs from other ensemble analysis techniques by the use of an adequate weighting of the input forecast models in order to obtain a combined estimation of meteorological parameters. Weights are calculated by least-square minimization of the differences between the model and the observed field during a so-called training period.

Although it can be applied successfully on continuous parameters like temperature, relative humidity, wind speed and mean sea level pressure, the Multimodel SuperEnsemble also gives good results when applied on the precipitation, a parameter quite difficult to handle with standard post-processing methods. Here we present a methodology for the Multimodel precipitation forecasts with a careful ensemble dressing via the precipitation PDF estimation.

1 Introduction

The Multimodel SuperEnsemble technique has been originally proposed by Krishnamurti et al. (1999) as a powerful statistical postprocessing method for a better estimation of weather forecast parameters with weights calculated in a training period. We have already applied it in Piemonte region (north-western Italy), a complex orographic area, to provide a more accurate forecast of several weather parameters (Cane and Milelli, 2006), including precipitation (Cane and Milelli, 2010). An exhaustive explanation of the Multimodel SuperEnsemble technique is provided in Cane and Milelli (2006) and Cane and Milelli (2010). Here we propose a probabilistic QPF evaluation with the use of a new

Multimodel SuperEnsemble dressing technique. This new approach, providing an estimation of the Probability Density Function (PDF) of precipitation, widens our knowledge of the precipitation field characteristics and is an effective support for operational weather forecast; it can also be used as an input for the hydrological forecast chain propagating the QPF uncertainty to the evaluation of its effects on the territory.

In Sect. 2 we describe the SuperEnsemble technique, while in Sect. 3 we compare the results of the calibrated ensemble with the original non-calibrated ensemble and with the SuperEnsemble probabilistic approach proposed by Stefanova and Krishnamurti (2002). Eventually, in Sect. 4 we draw some conclusion and outline the future perspectives.

2 Multimodel SuperEnsemble dressing

Arpa Piemonte manages a wide non-GTS weather station network. For this work we used the data collected in the period August 2004–April 2009 from 342 stations. The data were averaged over the 13 warning areas designed by Arpa Piemonte in collaboration with the Civil Protection Department (Fig. 1) on a 6-h basis up to +72 h and for each of them the maximum values of observed precipitation has been assigned. This spatial scale is practical for the use in an alert system on medium- and large-scale catchments and is the operational basis for the evaluation of danger levels and civil protection actions in Piemonte.

Each warning area contains on average 26 stations, with a minimum of 11 and a maximum of 39. In the analysis period, we have 96 460 values of both average and maximum 6-h precipitation, but more than 60% of them are cases of no precipitation.



Correspondence to: D. Cane
(daniele.cane@arpa.piemonte.it)

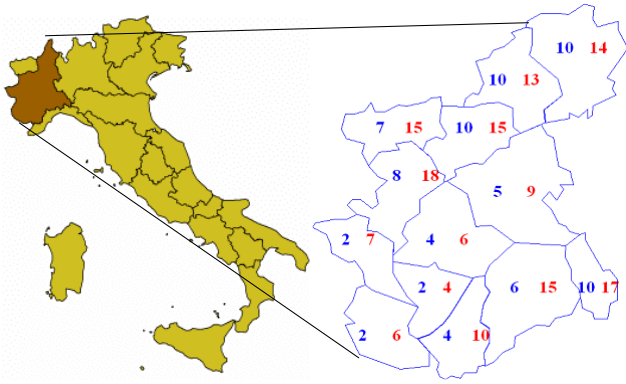


Fig. 1. Example of data observed in Piemonte region (north-western Italy) on 23 May 2007. Blue numbers represent the observed (or forecasted) 6-h precipitation (mm) averaged over the catchment, while red numbers are the maximum precipitation observed (or forecasted) in the catchment.

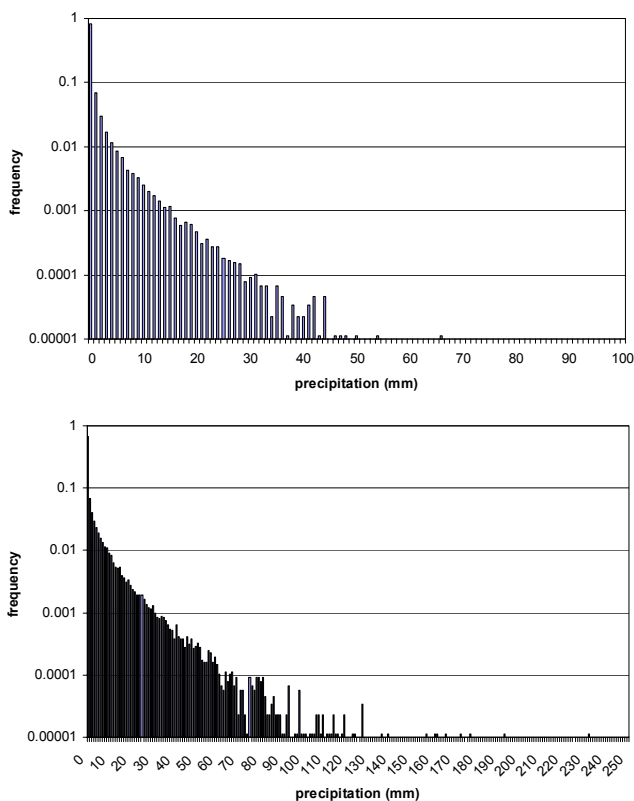


Fig. 2. Observed precipitation PDF: average values (left panel) and maximum values (right panel).

We first examined the observed PDF of these aggregated data, obtaining the results in Fig. 2.

For the use in an ensemble dressing the observed PDF however is not enough: we have to introduce a PDF conditioned to the forecasts. For each model we have to know which was the observed PDF given a single value forecast.

The models used in this research work are the ECMWF IFS global model (horizontal resolution: 0.25°) and the 0.0625° resolution limited area models of the COSMO Consortium encompassing North-Western Italy: COSMO-I7 (developed by Italian Air Force Weather Service, Arpa-Emilia Romagna, Arpa Piemonte), COSMO-EU (Deutscher Wetterdienst) and COSMO-7 (MeteoSwiss). It has to be highlighted that our operational implementation is based on the forecasts given by the ECMWF model and by the Italian version of the COSMO model only (see www.cosmo-model.org for a more comprehensive overview of the Consortium activities and developments), but for research purposes we have the possibility to also use the German and Swiss versions. The model forecasts are assigned to the same warning areas by taking the average and maximum values of the grid-points falling into the given area (ECMWF model: ~ 5 points per warning area; COSMO models: ~ 56 points per warning area).

For each model and for each run (00:00 and 12:00 UTC) we evaluated at first the best-fit PDF distribution with a confrontation among many distributions and we found the best agreement with the Weibull distribution (Weibull, 1951: blue curves in Fig. 3): this distribution is the only one fitting all the conditioned PDFs for different models.

For each run of the single/individual deterministic models we then evaluated the mean and the variance of the conditioned PDF of precipitation forecasts. In a Weibull PDF these two parameters are strongly correlated: we calculated the best fit mean-variance exponential relation (up to 40 mm/6 h for average values and up to 60 mm/6 h for maxima, with a $r^2 > 0.97$ for all the cases) and then we calculated the Weibull distribution parameters from the obtained fit curves, thus extrapolating the PDFs to higher precipitation rates where few data were available for a best fit (pink curves in Fig. 3).

We then weighted the obtained PDFs using weights obtained from the Brier scores evaluated in the training period (2 years before the forecast): for any given forecast time, for each available model the weights were assigned as the inverse of the Brier score normalized by the sum of the inverse of all the Brier scores.

Our dressing approach differs substantially from the Bayesian Model Averaging technique (see for example Raftery et al., 2005, McLean Slaughter et al, 2006) in the PDF evaluation method and in the Multimodel SuperEnsemble weighting calculation.

It differs also from the method proposed by Stefanova and Krishnamurti (2002): these authors derive a set of modified ensemble members directly from the Multimodel theory as

$$Fs_i = N a_i (F_i - \overline{F_i}) + \overline{O} \quad (1)$$

where N is the number of models, a_i are the SuperEnsemble weights, $(F_i - \overline{F_i})$ is the unbiased forecast, and \overline{O} is the observation mean in the training period. The so-obtained

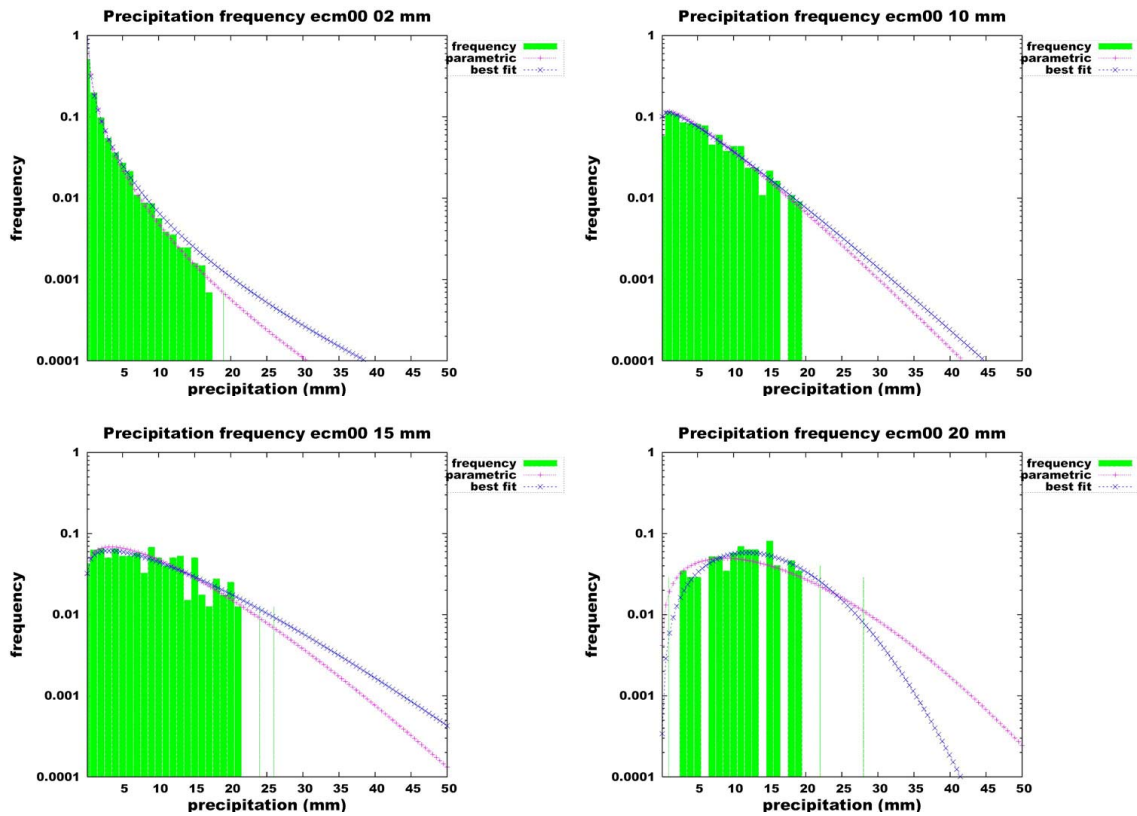


Fig. 3. Examples of conditioned PDF for the average precipitation over the warning areas (see Fig. 1) for ECMWF IFS run 00:00 UTC. From top left to bottom right: 2 mm, 10 mm, 15 mm, 25 mm forecasts. Pink line: Weibull distribution evaluated from distribution moments; blue line: best fit with a Weibull distribution (see text for details).

probabilistic ensemble members widen the ensemble but can bring to unphysical negative values for precipitation.

The ensembles are weighted by specific weights defined as follows:

$$w_i = \frac{c_i^k}{\sum_{i=1}^N c_i^k} \quad (2)$$

where c_i is the sum of the hit rate for the event and the hit rate for the non-event of the i^{th} model over the training period and k is an empirically chosen constant fixed, as best choice, at 3.

3 Results

We calculated the probabilistic ensemble from May 2007 to April 2009, using a sliding training period of two years before the forecast day and we verified our forecasts with the observations described in the previous section.

We compared our results with the non-calibrated ensemble (where the probabilities of passing a given threshold are simply the percentage of models of the ensemble exceeding it)

and with the probabilistic ensemble proposed by Stefanova and Krishnamurti (2002).

We evaluated a wide set of scores recommended by WMO (2008) for probabilistic QPF verification and in particular here we discuss Brier Skill Score and ROC (Relative Operating Characteristic) Area Skill Score. We evaluated the score errors with the bootstrapping technique proposed by Hamill (1999) using 1000 re-sampling subsets. The significance level chosen for the error evaluation was 95%.

In Figs. 4 and 5 (for average and maxima values, respectively) we show the skill scores for

- the non-calibrated ensemble,
- the Stefanova ensemble and for our Weibull-calibrated ensemble, using
- equal weights,
- the Stefanova weights and
- our weights obtained with the Brier score calculation.

For average values our ensemble is clearly more performing than the Stefanova and non-calibrated ensembles, both in terms of Brier Skill Score (with exception of the 10 mm

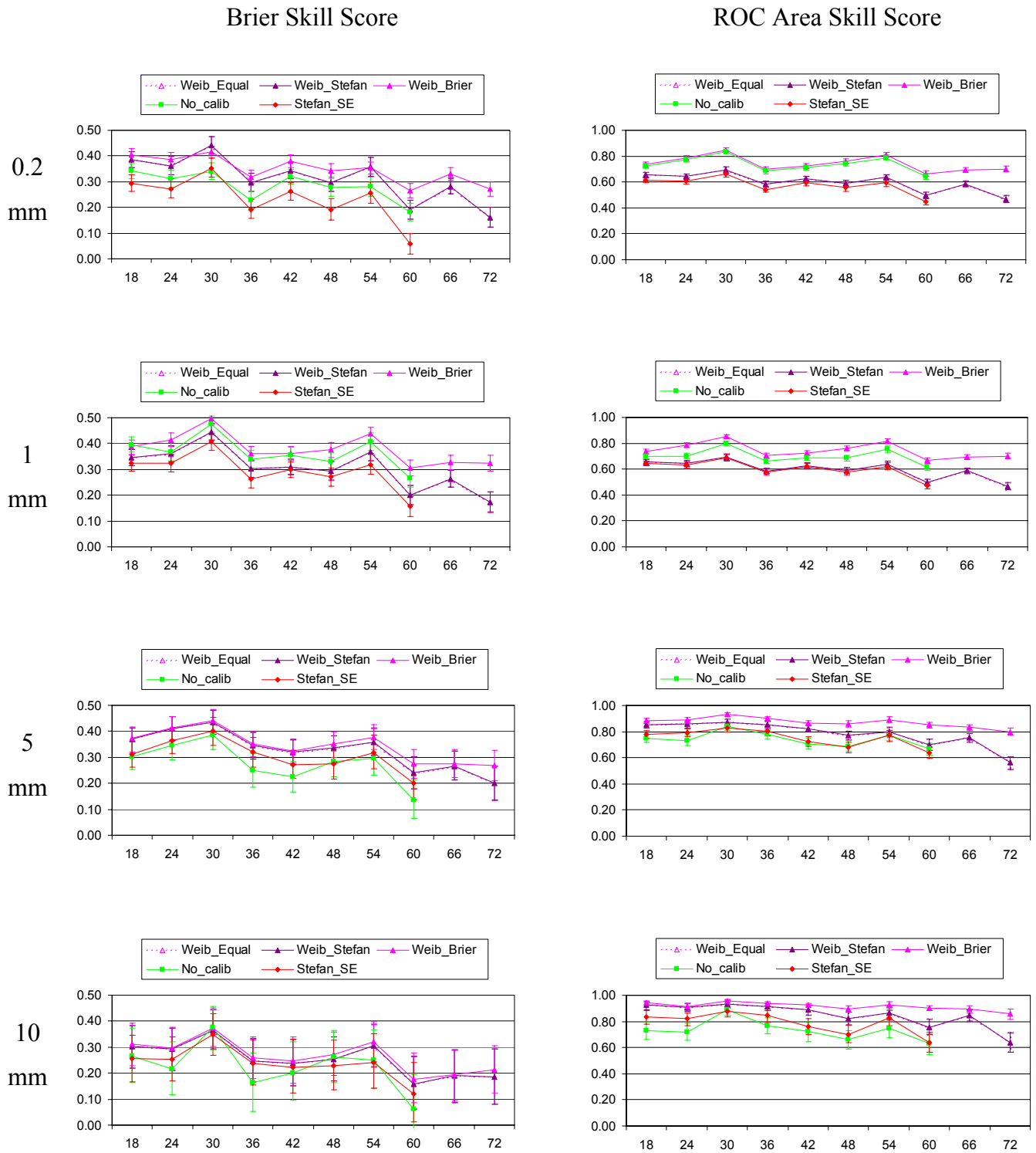


Fig. 4. Comparison of the Brier Skill Score (left-side panels) and the ROC Area Skill Score (right-side panels) results for the average precipitation values, calculated using the 0.2-, 1-, 5- and 10-mm thresholds, for the different ensembles: the Weibull-based approach with the Stefanova weights (purple line), the Brier score-based weights (pink line) and the equal weights (pink dashed line); the not-calibrated ensemble (green line); and the Stefanova ensemble (red line).

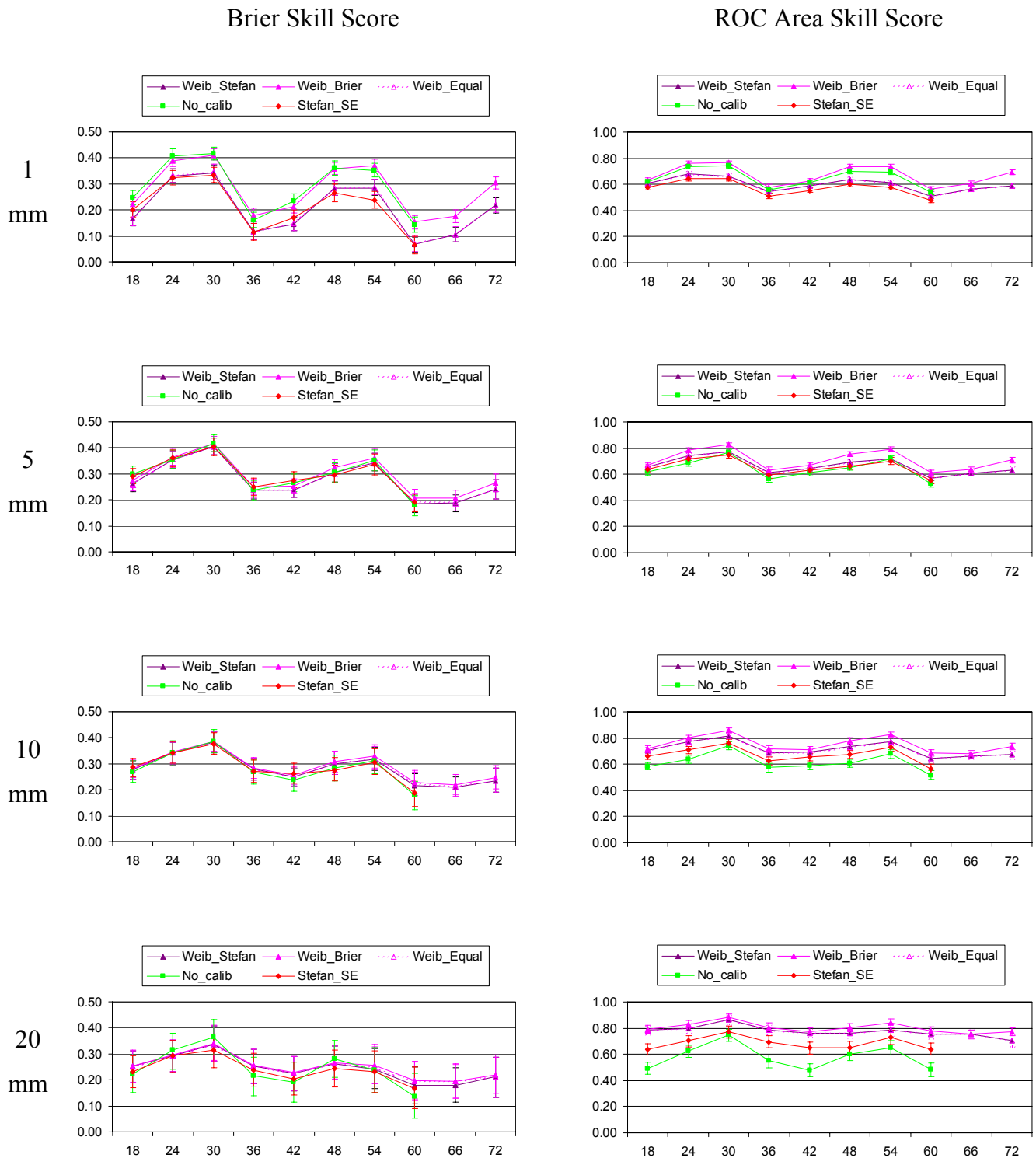


Fig. 5. Comparison of the Brier Skill Score (left-side panels) and the ROC Area Skill Score (right-side panels) results for the maximum precipitation values, calculated using the 1-, 5- and 10- and 20-mm thresholds, for the different ensembles: the Weibull-based approach with the Stefanova weights (purple line), the Brier score-based weights (pink line) and the equal weights (pink dashed line); the not-calibrated ensemble (green line); and the Stefanova ensemble (red line).

threshold, where it is still higher but not significantly) and of ROC Area Skill Score. The use of Stefanova weights with the Weibull distribution gives the same results as the use of equal weights (that is to say, it does not bring any improvement), and they are less performing than the Brier weights.

For maximum values the results are less clear: it is not possible to disentangle the various Brier Skill Scores, except for 1 mm threshold where the Weibull-calibrated ensembles and the non-calibrated ensemble over-perform the Stefanova ensemble and the equal weighted ensemble. The ROC Area Skill Score shows a significant prevalence of the Weibull-calibrated ensembles over the non-calibrated ensemble, the Stefanova ensemble and the equal-weighted Weibull ensemble. The ROC Area Skill Score is actually a measure of the predictive potential of the ensemble, that is to say, the ability of the ensemble to catch the forecast spread, therefore we have still room for improvement, but so far we did not find the correct way to weight the ensemble members to fulfil this potential in the case of maximum values.

4 Conclusions

We propose in this paper a new technique for ensemble dressing combining the observation PDFs conditioned to the forecasts and the Multimodel SuperEnsemble technique. The results are encouraging, with a clear improvement in forecasting the average precipitation over the warning areas, and a neutral effect in forecasting the maximum values. We are planning to apply this probabilistic precipitation forecasts as input of the Arpa Piemonte hydrological forecasting chain, in order to evaluate the uncertainties in the discharge calculations. However, the chosen spatial scale is practical for the use in an alert system of medium- and large-scale catchments, but it is too much coarse for the discharge calculations in the smallest catchments, where multimodel-based forecasts need instead to be provided at finer spatial resolution (e.g., at station location resolution).

Acknowledgements. This work is supported by the Italian Civil Defence Department. We thank the Deutscher Wetterdienst and MeteoSwiss for the use of their operational versions of the COSMO model in this research work. We are also grateful to our former colleague Marco Turco (now at the University of Barcelona) for the useful discussions.

Edited by: S. C. Michaelides

Reviewed by: two anonymous referees

References

- Cane, D. and Milelli, M.: Weather forecasts with Multimodel SuperEnsemble Technique in a complex orography region, *Meteorol. Z.*, 15(2), 1–8, 2006.
- Cane, D. and Milelli, M.: Multimodel SuperEnsemble technique for quantitative precipitation forecasts in Piemonte region, *Nat. Hazards Earth Syst. Sci.*, 10, 265–273, 2010, <http://www.nat-hazards-earth-syst-sci.net/10/265/2010/>.
- Krishnamurti T. N., Kishtawal, C. M., Larow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Improved weather and seasonal climate forecasts from Multimodel Superensemble, *Science*, 285, 1548–1550, 1999.
- Hamill, T. M.: Hypothesis tests for evaluating numerical precipitation forecasts, *Wea. Forecasting*, 14, 155–167, 1999.
- McLean Sloughter J., Raftery A. E. and Gneiting T.: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging, Tech. Rep. no. 496, Department of Statistics, University of Washington, 2006.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.* 133, 1155–1174, 2005.
- Stefanova L. and Krishnamurti T. N.: Interpretation of Seasonal Climate Forecast Using Brier Skill Score, The Florida State University Superensemble, and the AMIP-I Dataset, *J. Climate*, 15, 537–544, 2002.
- Weibull, W.: A statistical distribution function of wide applicability, *J. Appl. Mech.-Trans. ASME* 18(3), 293–297, 1951.
- WMO: Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models, published by WMO, 2008.